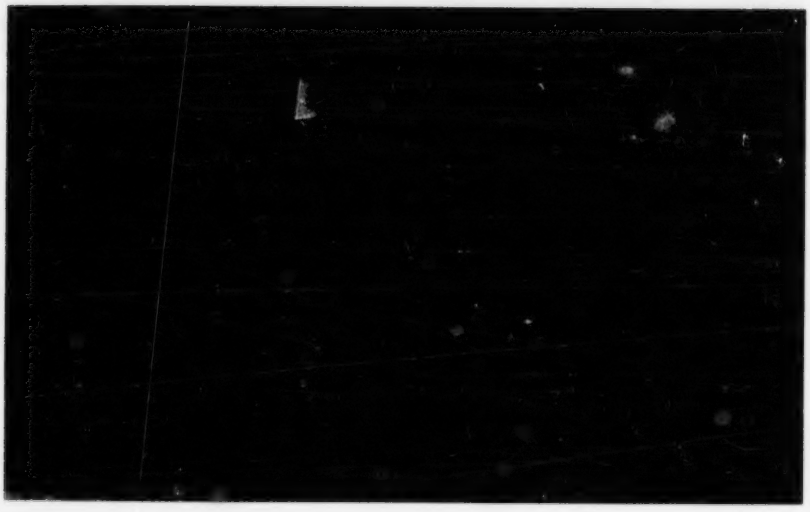
CENTRE DE RECHERCHE
SUR LES TRANSPORTS
(C.R.T.)
CENTRE FOR RESEARCH
ON TRANSPORTATION

UNIVERSITÉ DE MONTRÉAL
ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE POLYTECHNIQUE



CENTRE DE RECHERCHE SUR LES TRANSPORTS
UNIVERSITÉ DE MONTRÉAL
C.P. 6128, SUCCURSALE A,
MONTRÉAL
CANADA H3C 3J7

TÉLÉPHONE: (514) 343-7575
TÉLEX: (21) 05-561325
EMAIL: crt@crt.umontreal.ca
TÉLÉCOPIEUR (FAX): (514) 343-7121

91-05263/

CRT-776

Econometric Issues in Transportation Demand Forecasting Models: The Case of Pivot Point Modeling Procedure and the PODM Model

by

Bernard-André Genest[2]
Richard Laferrière[1]

[1] Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale A, Montréal, Québec, Canada H3C 3J7

[2] Département de sciences administratives, Université du Québec à Montréal, C.P. 6192, succursale A, Montréal, Québec, Canada H3C 4R2

Centre de recherche sur les transports — Publication #776
May 1991

Dépôt légal – 2ème trimestre 1991

RÉSUMÉ

Cette étude montre comment les résidus d'une analyse de régression sont introduits dans des modèles de prévision. Les implications économétriques de l'utilisation des résidus pour fins de prévision sont présentées. Deux modèles de prévision sont considérés: le modèle PODM et le modèle logit-pivot.

Mots-clés: modèle PODM, modèle économétrique de prévision

ABSTRACT

In this paper, we show how computed residuals from regression are introduced into forecasting models and discuss the econometric implications. We give two examples of such models: the first one is the PODM model and the second one is the incremental or pivot logit model.

Key-words: PODM model, econometric forecasting models

Table of Contents

1 INTRODUCTION 1

2 THE PODM MODEL 1
2.1 WELL SPECIFIED MODEL 1
2.2 MISSPECIFIED MODEL 4

3 THE PIVOT MODEL 7

4 CONCLUSION 8

5 REFERENCES 8

1 INTRODUCTION

In this paper, we show how computed residuals from regression are introduced into forecasting models and discuss the econometric implications. We give two examples of such models: the first one is the PODM model and the second one is the incremental or pivot logit model.

Forecasts of the air travel demand made by Transport Canada are partly based on the PODM model. We show how this model can be rewritten in order to explicit the role of the computed residuals into the forecasting model. We argue that computed residuals should not be included into the forecasting model since it increases unduely the variance of the forecast. This is the case wether the estimated model is well specified or misspecified.

Kumar (1980) has presented an incremental or pivot logit model. We show that this share model owns the same restrictions than those of the PODM model since it uses the computed residuals of the regression to do the forecastings.

2 THE PODM MODEL

2.1 WELL SPECIFIED MODEL

The PODM forecast model is defined as follows:

$$\frac{\hat{T}_{ij,f}}{T_{ij,b}} = \left(\frac{POP_{i,f}}{POP_{i,b}}\right)^{\hat{\beta}_1} \left(\frac{EMP_{j,f}}{EMP_{j,b}}\right)^{\hat{\beta}_2} \left(\frac{DIS_{ij,f}}{DIS_{ij,b}}\right)^{\hat{\beta}_3}, \tag{1}$$

where $T_{ij,b}$ is the observed total trips in the market ij during the base period; POP, EMP and DIJ stand for population, employment and distance, respectively; the subscripts b and f designate base and forecast period, respectively; β's are parameters and lastly, the "^" means that the variable or the parameter is an estimate. The estimation of the β's parameters comes from the following model:

$$T_{ij,b} = e^{\beta_0} POP_{i,b}^{\beta_1} EMP_{j,b}^{\beta_2} DIS_{ij,b}^{\beta_3} e^{\varepsilon_{ij}}. \tag{2}$$

The estimation of (2) requires linearization and is presented in (3).

$$\ln(T_{ij,b}) = \hat{\beta}_0 + \hat{\beta}_1 \ln(POP_{i,b}) + \hat{\beta}_2 \ln(EMP_{j,b}) + \hat{\beta}_3 \ln(DIS_{ij,b}) + \hat{\varepsilon}_{ij} \quad (3)$$

Let us now rewrite the PODM model as presented in (1) by using the logarithmic transformation and then by substituting $\ln(T_{ij,b})$ with its expression given in (3).

$$\begin{aligned} \ln(\hat{T}_{ij,f}) = & \quad \hat{\beta}_1 \ln(POP_{i,f}) + \hat{\beta}_2 \ln(EMP_{j,f}) + \hat{\beta}_3 \ln(DIS_{ij,f}) + \ln(T_{ij,b}) \\ & - \hat{\beta}_1 \ln(POP_{i,b}) - \hat{\beta}_2 \ln(EMP_{j,b}) - \hat{\beta}_3 \ln(DIS_{ij,b}) \\ = & \quad \hat{\beta}_1 \ln(POP_{i,f}) + \hat{\beta}_2 \ln(EMP_{j,f}) + \hat{\beta}_3 \ln(DIS_{ij,f}) + \hat{\beta}_0 + \hat{\varepsilon}_{ij} \end{aligned} \quad (4)$$

The last equation reveals that the PODM model includes the computed residual ($\hat{\varepsilon}_{ij}$) from the base period to forecast period f. We will now compare econometrically the PODM forecast with a model which omits the computed residuals in the forecasting. Hereafter it is referred as the best unbiased predictor (BUP) and defined as

$$\ln(\hat{T}^{*}_{ij,f}) = \hat{\beta}_1 \ln(POP_{i,f}) + \hat{\beta}_2 \ln(EMP_{j,f}) + \hat{\beta}_3 \ln(DIS_{ij,f}) + \hat{\beta}_0 . \quad (5)$$

The comparison is based on three criteria: point estimate, expected estimate, and variance of the estimate.

POINT ESTIMATES

For clarity purposes, the matrix notation is used in the proofs. Let's define N as the number of observations or number of markets in the sample, X a matrix with N rows and 4 columns: the first one being a constant, POP is the second, EMP and DIS are the others. The subscripts b and f will be defined as one specific market corresponding to the base period and the forecast period, respectively.

Let's start by comparing the forecast error implied in both procedures, the PODM (6), and the BUP (7).

$$\ln(T_f) - \ln(\hat{T}_f) = X_f(\beta - \hat{\beta}) + \varepsilon_f - \hat{\varepsilon}_b \quad (6)$$

$$\ln(T_f) - \ln(\hat{T}_f^*) = X_f(\beta - \hat{\beta}) + \varepsilon_f \quad (7)$$

The PODM approach compensates BUP forecast by adding the computed residual term ($\hat{\varepsilon}_b$). This means that if the calibrated model overestimates trips for some origin-destination pairs ($\hat{\varepsilon}_b < 0$), then the PODM forecast is smaller than the one of BUP.

The PODM forecast could be justified if random terms of the base period are autoregressive of order 1 with the random terms of the forecast period and the correlation is 1. This last assumption however violates the stationary assumption which states that the absolute value of the correlation coefficient must strictly be smaller than 1.

EXPECTED ESTIMATES

It is straightforward to prove that forecasting errors for both approaches are not biased, i.e. the expected value of (6) and (7) are both equal to zero. Therefore, even when comparing point estimates we identify systematic differences between the two approaches, both on average yield the same outcome.

VARIANCE ESTIMATES

Since both forecasting errors are unbiased, then the one with the smallest variance should be the privileged one. Below, we derive their respective variances. Before doing so, let us rewrite $\hat{T}_f$ and $\hat{T}_f^*$ as

$$\begin{aligned} \hat{T}_f &= (X_f M + (B - X_b M))T \\ &= (X_f M + C)T \\ \hat{T}_f^* &= X_f M T \end{aligned}$$

where $M = (X'X)^{-1}X'$

and B is a (1xN) vector equal to 0, except for the base period b which is equal to 1. The variance of the forecasting errors are equal to

$$V\{\ln(T_f)-\ln(\hat{T}_f)\} = E\{\varepsilon_f-(X_fM+C)\varepsilon\}^2$$

$$= \sigma^2(1+X_f(X'X)^{-1}X'_f+C'C) \quad (8)$$

$$V\{\ln(T_f)-\ln(\hat{T}^*_f)\} = E\{\varepsilon_f-X_fM\varepsilon\}^2$$

$$= \sigma^2(1+X_f(X'X)^{-1}X'_f) \quad (9)$$

Even though the PODM forecast is unbiased, it has a greater variance than that of the BUP. Thus, the inclusion of the computed residuals into the forecast computation is unsound since it unduely increases the variance.

The results presented so far would not surprise any econometrician since it is a well-known result that $\hat{T}^*_{ij,f}$, as defined in (2), is the best unbiased predictor.

2.2 MISSPECIFIED MODEL

One argument to justify the PODM forecast might be the presence of a systematic bias from the BUP forecast which could be related to the omission of relevant factors in the regression. Then forecasts could be improved with the inclusion of the computed residuals and therefore the missing variables. In Canada, for instance, omission to measure adequately cultural differences into a regression model induce overestimation of the number of trips between francophone and anglophone cities.

POINT ESTIMATES

Let us define the "real" model as:

$$T = X_0\beta_0+X_1\beta_1+\varepsilon\,. \quad (10)$$

Now let us assume that the relevant X_2 variables are not included into the regression model for some reason. The forecast errors in this case are given below.

$$T_f - \hat{T}_f = X_{0,f}\beta_0 + X_{1,f}\beta_1 + \varepsilon_f - (X_{0,f}\hat{\beta}_0 + \hat{\varepsilon}_b)$$

$$= (X_{0,f} - X_{0,b})(\beta_0 - \hat{\beta}_0) + (X_{1,f} - X_{1,b})\beta_1 + \varepsilon_f - \varepsilon_b \quad (11)$$

since $\hat{\varepsilon}_b = T_b - X_{0,b}\hat{\beta}_0$

$$= X_{0,b}(\beta_0 - \hat{\beta}_0) + X_{1,b}\beta_1 + \varepsilon_b \quad (12)$$

$$T_f - \hat{T}_f^* = X_{0,f}(\beta_0 - \hat{\beta}_0) + X_{1,f}\beta_1 + \varepsilon_f . \quad (13)$$

If explicative variables are constant ($X_{0,b} = X_{0,f}$, $X_{1,b} = X_{1,f}$) then the variation in the number of trips, according to (11), is solely produced by random variables ($\varepsilon_f - \varepsilon b$) as expected. However, (13) implies that change in the dependent variable, in this case, would be ($\varepsilon_f - \varepsilon_b + \hat{\varepsilon}_b$), which is different from the real change.

However, in general, when explicative variables do change from base period to forecast period, point estimates do not suggest in any way to include or to exclude computed residuals from the forecast.

EXPECTED ESTIMATES

It is a well-known result that the least squares estimates of the β's parameters are biased when relevant variables are omitted

$$\hat{\beta}_0 = (X'_0X_0)^{-1}X'_0T$$

$$= \beta_0 + (X'_0X_0)^{-1}X'_0X'_1\beta_1 + (X'_0X_0)^{-1}X'_0\varepsilon . \quad (14)$$

Let us now examine the expected estimates of the PODM and BUP forecasts in the case of a misspecified model. To do this, we take the expected value of (11) and (13).

$$E(T_f - \hat{T}_f) = (X_{0,f} - X_{0,b})(-(X'_0X_0)^{-1}X'_0X'_1\beta_1) + (X_{1,f} - X_{1,b})\beta_1 \quad (15)$$

$$E(T_f - \hat{T}_f^*) = (X_{0,f}) \quad (-(X'_0X_0)^{-1}X'_0X'_1\beta_1) + X_{1,f}\beta_1 \quad (16)$$

The only justification for the PODM forecasting may be found in (15). If one assumes that included variables are orthogonal to the missing variables ($X'_0X_1 = 0$) and that excluded variables are the same in the base period and the forecasting period ($X_{f,1} = X_{b,1}$) then the PODM forecast is unbiased. Otherwise, it is inconclusive whether one approach should be preferred.

VARIANCE ESTIMATES

Since point estimates and expected estimates do not lead to any support to either PODM forecasting or BUP forecasting then one might be inclined to choose the procedure that would yield the smallest variance. Below, we derive the variances of the two forecasting procedures. Before to do so let's rewrite $\hat{T}_f$ and $\hat{T}_f^*$ as

$$\begin{aligned}\hat{T}_f &= X_{0,f}\beta_0 + T_b - X_{0,f}\beta_0 = (X_{0,f}M_0 + (B - X_{0,b}M_0))T \\ &= (X_{0,f}M_0 + C_0)T \qquad (17)\end{aligned}$$

$$\hat{T}_f^* = X_{0,f}\beta_0 = (X_{0,f}M_0)T \qquad (18)$$

The variance of the forecasting errors are equal to

$$\begin{aligned}V\{\ln(T_f) - \ln(\hat{T}_f)\} &= E\{\varepsilon_f - (X_{0,f}M_0 + C_0)\varepsilon\}^2 \\ &= \sigma^2(1 + X_{0,f}(X_{0,b}'X_{0,b})^{-1}X'_{0,f} + C_0'C_0) \qquad (19)\end{aligned}$$

$$\begin{aligned}V\{\ln(T_f) - \ln(\hat{T}_f^*)\} &= E\{\varepsilon_f - X_{0,f}M_0\varepsilon\}^2 \\ &= \sigma^2(1 + X_{0,f}(X_{0,b}'X_{0,b})^{-1}X'_{0,f}) \qquad (20)\end{aligned}$$

Therefore the variance of the PODM forecast error is greater than the BUP one.

3 THE PIVOT MODEL

The forecast pivot model is defined as follows

$$\hat{p}^m_{ij,f} = \frac{p^m_{ij,b} \quad \exp(\hat{V}^m_{ij,f} - \hat{V}^m_{ij,b})}{\sum_l p^l_{ij,b} \quad \exp(\hat{V}^l_{ij,f} - \hat{V}^l_{ij,b})}, \tag{21}$$

where $\hat{p}^m_{ij,f}$ is the forecasted share of the mode m for the period f, $p^m_{ij,b}$ is the observed share of the mode m during the period b, $\hat{V}^m_{ij,b}$ and $\hat{V}^m_{ij,b}$ are, respectively, the computed service level of the mode m during the base and forecast period.

The estimated logit can be represented as follows

$$p^m_{ij,b} = \frac{\exp(\hat{V}^m_{ij,b} + \hat{\varepsilon}^m_{ij,b})}{\sum_l \exp(\hat{V}^l_{ij,b} + \hat{\varepsilon}^l_{ij,b})}, \tag{22}$$

where $\hat{\varepsilon}^m_{ij,b}$ is the computed residual from the base period.

We now rewrite the pivot model by using the logarithmic transformation and then by substituting $p^m_{ij,b}$ with its expression given in (22).

$$\begin{aligned}
\ln\left(\frac{\hat{p}^m_{ij,f}}{\hat{p}^l_{ij,f}}\right) &= \ln\left(\frac{p^m_{ij,b} \quad \exp(\hat{V}^m_{ij,f} - \hat{V}^m_{ij,b})}{p^l_{ij,b} \quad \exp(\hat{V}^l_{ij,f} - \hat{V}^l_{ij,b})}\right) \\
&= \ln(p^m_{ij,b}) + \hat{V}^m_{ij,f} - \hat{V}^m_{ij,b} - \ln(p^l_{ij,b}) - \hat{V}^l_{ij,f} + \hat{V}^l_{ij,b} \\
&= \hat{V}^m_{ij,f} - \hat{V}^l_{ij,f} + \hat{\varepsilon}^m_{ij,b} - \hat{\varepsilon}^l_{ij,b},
\end{aligned} \tag{23}$$

It is obvious now that the forecast done by the pivot model such as presented in (23) has the same restrictions than those of the PODM. Indeed the model (23) has the same linear form (24) that was discussed in the preceding section.

$$\ln(\hat{T}_f) = X_f\beta + \hat{\varepsilon}_b \,, \qquad (21)$$

where $\hat{T}_f = \ln\left(\frac{\hat{p}^m_{ij,f}}{\hat{p}^l_{ij,f}}\right)$

$$X_f\beta = \hat{V}^m_{ij,f} - \hat{V}^l_{ij,f}$$

$$\hat{\varepsilon}_b = \hat{\varepsilon}^m_{ij,b} - \hat{\varepsilon}^l_{ij,b}$$

Therefore the pivot model applied to aggregate share yields unbiased forecast as the logit model but with a greater variance.

4 CONCLUSION

In this paper we have shown that the two forecasting models, namely the PODM and the pivot model, differ from the traditional one by using the computed residuals from the regression. This increases unduely the variance of the forecast produced by these two models.

5 REFERENCES

KUMAR, A., "Pivot Point Modelling Procedures in Demand Estimation", **Transportation Engineering Journal of ASCE**, v.106, no.TE6, November 1981.

Transport Canada, Passenger Origin-Destination Model, TP 3215, June 1981.